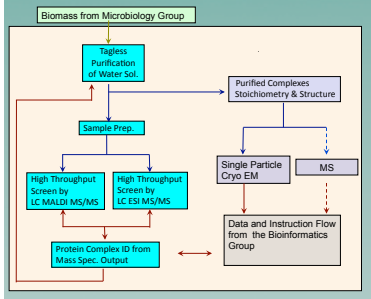
High Throughput Identification, Purification and Structural Characterization of Water Soluble Protein Complexes in *Desulfovibrio vulgaris*

Ming Dong[1], Bong-Gyoon Han[1], Haichuan Liu[2], Maxim Shatsky[1,3], Steven E. Brenner[1,3], Pablo Arbelaez[3], Jitendra Malik[3], Dieter Typke[1], Terry C. Hazen[1], Jil T. Geller[1], Harry J. Sterling[3], Lee Yang[1], Megan Choi[1], Evelin D. Szakal[2], Simon Allen[2], Steven C. Hall[2], Susan J. Fisher[1,2], Evan R. Williams[3], John-Marc Chandonia[1], Jian Jin[1], H. Ewa Witkowska[2], Robert M. Glaeser[1], Mark D. Biggin[1].

Ecosystems and Networks Integrated with Genes and Molecular Assemblies [1] Lawrence Berkeley National Laboratory, Berkeley CA; [2] University of California, San Francisco CA; [3] University of California, Berkeley CA

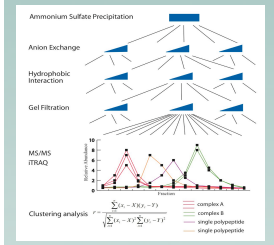
VIMSS Virtual Institute for Microbial Stress and Survival — http://vimss.lbl.gov/

PROTEIN COMPLEX IDENTIFICATION



The ENIGMA Protein Complex Analysis Project (PCAP) has two major goals: **1**. to develop an integrated set of high throughput pipelines to identify and characterize multi-protein complexes in a microbe more swiftly and comprehensively than currently possible, and **2.** to use these pipelines to elucidate and model the protein interaction networks regulating stress responses in *Desulfovibrio vulgaris*.

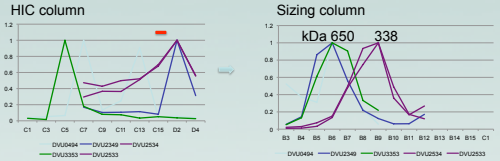
Work flow. The yellow arrow represents transfer of microbial cells, the purple arrows the transfer of protein samples, and the red arrows represent the transfer of data or instructional control functions between groups. The dotted arrows indicate a new pipeline, native ESI MS, under development.



Tagless Strategy Concept

Our scheme for the tagless purification of water soluble complexes. 10 g of protein from a crude bacterial extract is first fractionated by ammonium sulfate precipitation and then by a series of chromatographic steps: anion exchange (IEX), hydrophobic interaction (HIC), and finally size exclusion (Gel Filtration). Fractions from the last chromatography step are trypsin digested and peptides labeled with iTRAQ reagents to allow multiplexing and quantitation during mass spectrometric analysis. Elution profiles of identified proteins are then subjected to clustering analysis.

Multidimensional Tagless Strategy

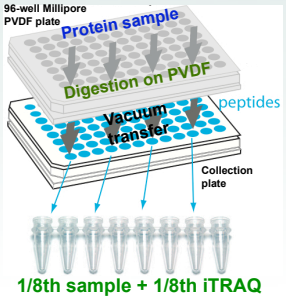


Monomer MW:
DVU0494: 41 kDa DVU2349: 97 kDa DVU3353: 43 kDa DVU2534: 38 kDa DVU2533: 87 kDa

Identified complexes:
Heteromultmer: DVU2533, DVU2534 (Phenylalanyl-tRNA synthetase) Homomultimer: DVU0494, DVU2349, DVU3353
The red bar fraction from HIC column was further fractionated by sizing column.

When eluting protein complexes from a column, some proteins that do not belong to the same complex could still co-migrate together by chance. By using data from two orthogonal columns, the chance of reducing such false positives and finding authentic complexes will be increased.

On PVDF Membrane digestion and Miniaturization of iTRAQ Labeling



- Protein ID scores: PVDF≈Solution
- Hydrophobicity of identified peptides: PVDF≈Solution (only soluble proteins were studied)
- Accuracy of relative quantitation: PVDF≈Solution
- Precision of relative quantitation: PVDF≤Solution

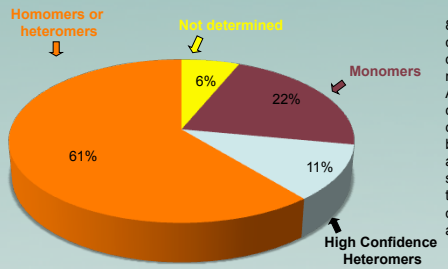
Features:

- ✓ High throughput
- ✓ Amenable to automation
- ✓ Savings in reagent cost

Progress and target of Tagless Strategy:

Fraction	1D (sizing direction)	2D (HIC direction)	target
IEX	21	10	28
HIC	221	102	298
SEC	4,194	NA	5662
%space	75%	32%	100%

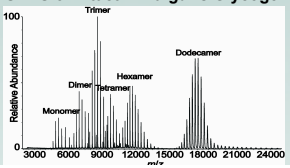
*Dv*H Protein Complexes Identified To Date



887 proteins have been confidently identified to date. Of these ~72% migrate as multimers. Among these, 45 high confidence heteromeric complexes were identified by additional manual analysis. Detailed statistical analysis of the tagless data is underway to define specific complexes automatically.

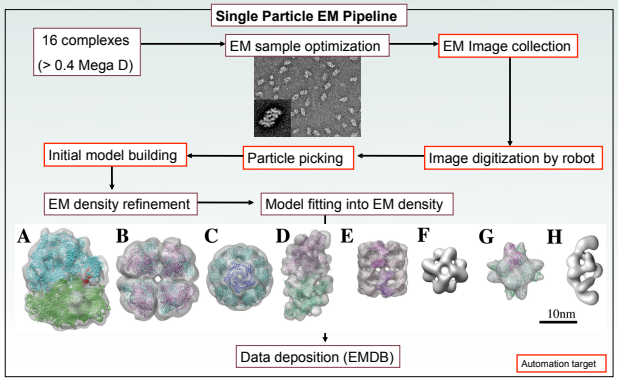
PROTEIN COMPLEX VALIDATION

ESI MS of Intact *D. vulgaris* Glycogen Phosphorylase Dodecamer



Glycogen Phosphorylase	Amino acid sequence Molecular Weight (Da)	Experimental Molecular Weight (Da)	Deviation (%)
Monomer	97,358	97,630 ± 7	+ 0.28
Dimer	194,716	195,254 ± 8	+ 0.28
Trimer	292,074	292,499 ± 54	+ 0.15
Tetramer	389,432	390,220 ± 118	+ 0.20
Hexamer	584,147	585,053 ± 196	+ 0.16
Dodecamer	1,168,295	1,154,009 ± 167	- 1.23

Through collaboration with Sterling and Williams at UC Berkeley, we analyzed a few protein complexes purified from the tagless strategy using electrospray ionization mass spectrometry (ESI MS). In the ESI MS shown above, the expected dodecamer is accompanied by smaller subunit assemblies: hexamer, tetramer, trimer and dimer. At the same time, other theoretically possible homomers (e.g. penta-, hepta, octa-mer) are missing, pointing to a non-random distribution of species. These lower mass assemblies originated in solution and might be artifacts of sample handling – further studies are needed to unequivocally establish their origin.

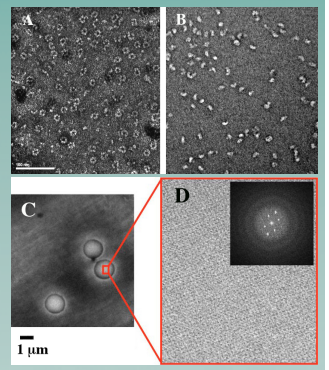


EM HIGH THROUGHPUT PIPELINE DEVELOPMENT

Robust Sample-Preparation Technology

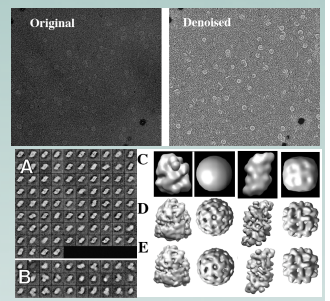
Half of the particles were not solved in high throughput mode. Some particles may be intrinsically flexible and structurally heterogeneous [possibly our novel ring-like assembly of a carbohydrate phosphorylase (A)]. Others seem likely to have been damaged by adsorption to the carbon film [possibly Proline dehydrogenase (B)].

We are testing whether proteins that are immobilized on support films with surface chemistries similar to chromatography matrix-materials retain a more homogenous structure, both in negative stain and as cryo-EM specimens. Monolayers with defined surface-chemistries can be picked up by holey EM grid (C). Ni-NTA derivatized lipid monolayers and streptavidin 2D crystal layers grown on top of lipid monolayers (D) are being used for complexes with His or Strep tags, respectively. As other alternative, DNA-derivatized surface layer is being developed.



Automation of Particle Picking

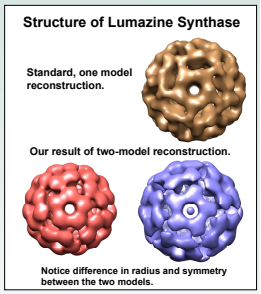
Particle picking from EM images is a time consuming process, occupying more than 50% of human time during EM structural work. We have developed an automated particle picking system called TextonSVM. Our approach models the appearance of molecular structures based on their texture and separates them from the background, applicable in data-driven denoising purposes (right top). The resolution of reconstructions obtained with the particles picked by TextonSVM is comparable to the one obtained with manually selected particles, at a fraction of the time and effort for the user (Fig. A, B, C, D, E).



(A) Examples of good classes. (B) Outliers. (C) Starting initial models. (D) Refined structures from automatically picked particles, polished further by classification. (E) Refined structures from manually picked particles.

Automated Multi-model Reconstruction

Biological macromolecules can adopt multiple conformational and compositional states due to structural flexibility and alternative subunit assemblies. This structural heterogeneity poses a major challenge in the study of macromolecular structure using single particle electron microscopy. We have developed a fully automated, unsupervised method for the three-dimensional reconstruction of multiple structural models from heterogeneous data. We tested our method on one synthetic and three distinct experimental datasets. The tests include the cases where a macromolecular complex exhibits structural flexibility and cases where a molecule is found in ligand-bound and unbound states. We propose the use of our approach as an efficient way to reconstruct distinct multiple models from heterogeneous data.



Reference:

Han, BG et al. Proc Natl Acad Sci U S A. 2009;106(39):16580-5.

Dong, M et al. J Proteome Res. 2008; 7(5):1836-49.

ACKNOWLEDGEMENTS

ENIGMA is a Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.